Panel Data Analyais I

Ye Wang
University of North Carolina at Chapel Hill

Linear Methods in Causal Inference
POLI784

Review

- We reviewed several variants of RDD in the previous class.
- We have fuzzy RDD when the treatment is not a deterministic function of the running variable.
- It can be analyzed as in the setting with instrumental variables.
- When the growth rate of the outcome and the treatment changes across the cutoff, we have the kink design.
- The running variable could be multi-dimensional or discrete.
- It is possible to generalize the RDD estimate when there are multiple cutoffs.

What is unique about panel data?

- In a typical panel dataset, each unit $i \in \{1, 2, \ldots, N\}$ is observed for T periods.
- We know the outcome Y_{it}, treatment status D_{it}, and some covariates $\mathbf{X}_{it}$.
- We use $\mathbf{U}_i$ to denote unobservable time-invariant confounders.
- We use superscript to denote the history of a variable: $\mathbf{Y}_i^{s:t}$.
- When T is short and N is large, we call it panel data; when T is long and N is moderate, we call it time-series cross-sectional (TSCS) data or long panel data.
- Household survey over three years vs. country-level data over fifty years.
- In panel data, asymptotics relies on a large N; in TSCS data, both N and T grow to infinity.
- It is known as longitudinal data in other disciplines.

What is unique about panel data?

- Why don't we just use the classical estimators (regression, weighting, AIPW, etc.)?
  - The dynamic structure allows us to relax the identification assumption,
  - SUTVA might be violated,
  - Observations are dependent.
- In panel data, SUTVA means

$$Y_{it} = \begin{cases} Y_{it}(0), \ D_{it} = 0 \\ Y_{it}(1), \ D_{it} = 1. \end{cases}$$

- The individualistic treatment effect $\tau_{it} = Y_{it}(1) - Y_{it}(0)$.
- SUTVA excludes the existence of anticipation or dynamic treatment effect: $Y_{it} = Y_{it}(D_{it}) = Y_{it}(\mathbf{D}_i^{1:T})$.
- It implies that $\mathbf{D}_i^{1:(t-1)}$ will not be confounders.
- Remember that treatment effect heterogeneity means $\tau_{it} = \tau_t(\mathbf{X}_i^{1:t}, \mathbf{U}_i)$.

Identification in panel data

- In the cross-sectional setting, we need unconfoundedness:

$$D_i \perp \{Y_i(0), Y_i(1)\} | \mathbf{X}_i.$$

- In panel data, we observe the history of each variable, hence the weakest assumption will be

$$D_{it} \perp \{Y_{it}(0), Y_{it}(1)\} | \mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t}, \mathbf{U}_i.$$

- It is too weak for identification.
- In practice, people strengthen the assumption along two different directions.

Identification in panel data

- Sequential ignorability:

$$D_{it} \perp \{Y_{it}(0), Y_{it}(1)\} | \mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t}.$$

- It prevents unobservable confounders from affecting treatment assignment: $P(D_{it} = 1) = g_t(\mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t})$.
- Strict exogeneity:

$$D_{it} \perp \{Y_{is}(0), Y_{is}(1)\} | \mathbf{X}_i^{1:t}, \mathbf{U}_i,$$

- It prevents the outcome history from affecting treatment assignment: $P(D_{it} = 1) = g_t(\mathbf{X}_i^{1:t}, \mathbf{U}_i)$.
- We always require $\varepsilon < g_t(\cdot) < 1 - \varepsilon$.

Ideal experiment behind the assumptions

- The two assumptions are based upon two different ideal experiments.
- Under sequential ignorability, the experimenter adjusts the probability of being treated for any unit dynamically based on the observed outcome.
- On Feb. 14, your probability of being vaccinated is 0.72 if you are an old Asian male who have not been infected by Covid.
- Data available to the analyst include each unit's observable attributes, health status (outcome), and treatment status over time.
- The experimenter observes the same data.

Ideal experiment behind the assumptions

- Under strict exogeneity, the experimenter knows all the unobservable attributes and specifies $g_t(\cdot)$ in a "pre-analysis plan" without conditioning on the outcome.
- It is known as "baseline randomization".
- On Feb. 14, your probability of being vaccinated is 0.72 if you are an old Asian male who loves tequila.
- Data available to the analyst do not include the unobservable attributes.
- The experimenter possesses more information than the analyst.

Ideal experiments behind the two assumptions

- Under sequential ignorability, the analyst observes all the variables that may affect treatment assignment.
- The remaining task is to infer the probability of being treated.
- All the methods we have learned can still be applied with some modifications.
- Under strict exogeneity, we have the problem of omitted variables as some confounders ($\mathbf{U}_i$) are unobservable.
- In this case, it is usually more challenging to infer the treatment assignment mechanism than to model the outcome variable.
- The outcome has a larger variation, which allows us to test the validity of the outcome model.

Estimation under strict exogeneity

- Note that the strict exogeneity assumption is justified by the following outcome model:

$$Y_{it} = m_t(D_{it}, \mathbf{X}_i^{1:t}, \mathbf{U}_i) + \varepsilon_{it}$$
$$E[\varepsilon_{is}|D_{it}, \mathbf{X}_{it}, \mathbf{U}_i] = 0,$$

  which is still too general for identification.
- In practice, we impose structural restrictions to simplify the model.
- Only contemporary values of $\mathbf{X}$ are confounders:

$$Y_{it} = m_t(D_{it}, \mathbf{X}_{it}, \mathbf{U}_i) + \varepsilon_{it}.$$

- The effects of $\mathbf{X}$ and $\mathbf{U}$ are additive:

$$Y_{it} = \tau_{it} D_{it} + f_t(\mathbf{X}_{it}) + h_t(\mathbf{U}_i) + \varepsilon_{it}.$$

- $\mathbf{X}$ affect Y in a linear manner and $h_t(\mathbf{U}_i)$ has a low-dimensional representation.

Estimation under strict exogeneity

- For example, we can assume that $h_t(\mathbf{U}_i) = \mu + \alpha_i + \xi_t$, then

$$Y_{it} = \mu + \tau_{it} D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \varepsilon_{it}.$$

- This is the two-way fixed effects (TWFE) model with heterogeneous treatment effects.
- α_i and ξ_t are known as unit and period fixed effects.
- Now, the assumption of strict exogeneity becomes: $E[\varepsilon_{is} | D_{it}, \mathbf{X}_{it}, \alpha_i, \xi_t] = 0$ for any s.
- The classical TWFE model further assumes that the treatment effect is homogeneous:

$$Y_{it} = \mu + \tau D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \varepsilon_{it}.$$

Estimation of the TWFE model

- Suppose we know the values of α_i and ξ_t, then we can estimate τ and β with OLS as in classic regression.
- But neither is known in practice.
- We have more than one observation of each unit and each period, hence eliminating α_i and ξ_t becomes possible.
- We need to impose two extra conditions for identification:

$$\sum_{t=1}^{T} \xi_t = 0, \sum_{i=1}^{N} \alpha_i = 0.$$

- These conditions specify the reference point of α_i and ξ_t and are not unique.

Estimation of the two-way fixed effects model

- For any random variable Y_{it}, let's define

$$\bar{Y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} Y_{it},\ \bar{Y}_{.t} = \frac{1}{N}\sum_{i=1}^{N} Y_{it},\ \text{and}\ \bar{Y} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it}$$

- Note that

$$\bar{Y}_{i.} = \mu + \tau\bar{D}_{i.} + \bar{\mathbf{X}}_{i.}\beta + \alpha_i + \bar{\varepsilon}_{i.}$$

- We subtract the equation above from the TWFE model, and obtain

$$Y_{it} - \bar{Y}_{i.} = \tau(D_{it} - \bar{D}_{i.}) + (\mathbf{X}_{it} - \bar{\mathbf{X}}_{i.})\beta + \xi_t + \varepsilon_{it} - \bar{\varepsilon}_{i.}$$

- We have eliminated α_i from the outcome model.

Estimation of the TWFE model

- Similarly, we have

$$\bar{Y}_{.t} = \mu + \tau \bar{D}_{.t} + \bar{\mathbf{X}}_{.t}\beta + \xi_t + \bar{\varepsilon}_{.t}$$

- Subtracting it from the previous equation, we have

$$\begin{aligned} Y_{it} - \bar{Y}_{i.} - \bar{Y}_{.t} = &-\mu + \tau(D_{it} - \bar{D}_{i.} - \bar{D}_{.t}) \\ &+ (\mathbf{X}_{it} - \bar{\mathbf{X}}_{i.} - \bar{\mathbf{X}}_{.t})\beta + \varepsilon_{it} - \bar{\varepsilon}_{i.} - \bar{\varepsilon}_{.t} \end{aligned}$$

- This looks like a classical regression regression except for $-\mu$.

Estimation of the TWFE model

- To eliminate $-\mu$, note that

$$\bar{Y} = \mu + \tau \bar{D} + \bar{\mathbf{X}}\beta + \bar{\varepsilon}.$$

- We add this equation back to the previous one, and obtain

$$\begin{aligned} Y_{it} - \bar{Y}_{i.} - \bar{Y}_{.t} + \bar{Y} &= \tau(D_{it} - \bar{D}_{i.} - \bar{D}_{.t} + \bar{D}) \\ &+ (\mathbf{X}_{it} - \bar{\mathbf{X}}_{i.} - \bar{\mathbf{X}}_{.t} + \bar{\mathbf{X}})\beta + \varepsilon_{it} - \bar{\varepsilon}_{i.} - \bar{\varepsilon}_{.t} + \bar{\varepsilon}. \end{aligned}$$

Estimation of the TWFE model

- Define
$$\tilde{Y}_{it} = Y_{it} - \bar{Y}_{i.} - \bar{Y}_{.t} + \bar{Y}$$
$$\tilde{D}_{it} = D_{it} - \bar{D}_{i.} - \bar{D}_{.t} + \bar{D}$$
$$\tilde{\mathbf{X}}_{it} = \mathbf{X}_{it} - \bar{\mathbf{X}}_{i.} - \bar{\mathbf{X}}_{.t} + \bar{\mathbf{X}}$$
$$\tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_{i.} - \bar{\varepsilon}_{.t} + \bar{\varepsilon}.$$
- Then the previous equation can be written as
$$\tilde{Y}_{it} = \tau \tilde{D}_{it} + \tilde{\mathbf{X}}_{it}\beta + \tilde{\varepsilon}_{it}.$$
- Note that $E[\tilde{\varepsilon}_{it} | \tilde{D}_{it}, \tilde{\mathbf{X}}_{it}] = 0$ due to strict exogeneity.
- Both τ and β can be estimated via OLS.
- This is known as the within estimator for the TWFE model.

Inference of the TWFE model

- In panel data, it is common to assume that the error terms are correlated within units (over periods) but not between units.
- The variance of $\begin{pmatrix} \hat{\tau} \\ \hat{\beta} \end{pmatrix}$ takes the familiar sandwich form:

$$Var \begin{pmatrix} \hat{\tau} \\ \hat{\beta} \end{pmatrix} = (\mathbf{X}^{\dagger\prime}\mathbf{X}^{\dagger})^{-1}(\mathbf{X}^{\dagger}\tilde{\varepsilon}\tilde{\varepsilon}'\mathbf{X}^{\dagger\prime})(\mathbf{X}^{\dagger\prime}\mathbf{X}^{\dagger})^{-1},$$

  where $\mathbf{X}^{\dagger} = (\tilde{D}, \tilde{\mathbf{X}})$.
- The variance can be estimated by either some heteroscedasticity and auto-correlation (HAC) consistent variance estimator or block bootstrap.
- In practice, blocks may differ from units (e.g., provinces vs. individuals).

Inference of the TWFE model

- Moreover, as $N \to \infty$

$$\frac{\hat{\tau} - \tau}{\sqrt{Var(\hat{\tau})}} \to N(0, 1)$$

  if the correlation between the random errors is weak.
- Therefore, we can easily construct the 95% confidence interval for $\hat{\tau}$.
- In block bootstrap, we resample the units rather than the observations.

TWFE models: application

- We use the study in Hainmueller and Hangartner (2019) for illustration.
- They studied the impacts of indirect democracy on naturalization of immigrants in Swiss municipalities.
- There are $1,211$ municipalities over 19 years.
- The treatment indicator equals 1 if the municipality relies on elected officials rather than popular referendums for naturalization decisions.
- The outcome is naturalization rate of municipality i in year t.

TWFE models: application

- Conventionally, we can estimate the model via the package *plm* in R.

```
## The TWFE estimate is 1.339325
```

```
## The SE estimate is 0.1863711
```

TWFE models: application

- A more modern approach is to use the *fixest* package.

```
## OLS estimation, Dep. Var.: nat_rate_ord
## Observations: 22,971
## Fixed-effects: bfs: 1,209,  year: 19
## Standard-errors: Clustered (bfs)
##           Estimate Std. Error t value   Pr(>|t|)
## indirect  1.33932   0.186525 7.18039 1.2117e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
## RMSE: 4.09541     Adj. R2: 0.152719
##                 Within R2: 0.005173
```

Caveats of the TWFE models

- Note how many assumptions we need for the model to work!
  - SUTVA,
  - Strict exogeneity,
  - Correct model specification,
  - Homogeneous treatment effect.
- Suppose the first three assumptions are satisfied but the treatment effects are heterogeneous.
- Following the same logic in Aronow and Samii (2016), we can show that

$$\hat{\tau} \rightarrow \sum_{D_{it}=1} w_{it}\tau_{it}, \text{with } \sum_{D_{it}=1} w_{it} = 1.$$

- Even worse, some w_{it} can be negative (Chaisemartin and D'Haultfœuille 2020).
- It means that $\hat{\tau}$ is not a convex combination of τ_{it}.
- $\hat{\tau}$ may not be representative of τ_{it} at all.

Caveats of the TWFE models

- In reality, SUTVA is often violated as dynamic treatment effects (or carryover) are common (Imai and Kim 2019).
- Treatment assignment can be affected by both the unobservable confounders and the outcome history (feedback).
- $h_t(\mathbf{U}_i)$ can be more complicated than $\mu + \alpha_i + \xi_t$.
- We say that treatment assignment follows the structure of staggered adoption if $D_{it} = 1$, then $D_{is} = 1$ for any $s > t$.
- Once a unit is treated, it will always be under treatment.
- Many caveats are avoided under staggered adoption.

References I

Aronow, Peter M, and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60 (1): 250–67.

Chaisemartin, Clément de, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*.

Hainmueller, Jens, and Dominik Hangartner. 2019. "Does Direct Democracy Hurt Immigrant Minorities? Evidence from Naturalization Decisions in Switzerland." *American Journal of Political Science* 63 (3): 530–47.

Imai, Kosuke, and In Song Kim. 2019. "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63 (2): 467–90.